

On the ADI method for the Sylvester Equation and the optimal- \mathcal{H}_2 points

Garret M. Flaggs and Serkan Gugercin

Department of Mathematics, Virginia Tech.
Blacksburg, VA, 24061-0123, USA
{garretf,gugercin}@vt.edu

Preprint submitted to *Applied Numerical Mathematics*, January 23, 2012

Abstract

The ADI iteration is closely related to the rational Krylov projection methods for constructing low rank approximations to the solution of Sylvester equation. In this paper we show that the ADI and rational Krylov approximations are in fact equivalent when a special choice of shifts are employed in both methods. We will call these shifts pseudo \mathcal{H}_2 -optimal shifts. These shifts are also optimal in the sense that for the Lyapunov equation, they yield a residual which is orthogonal to the rational Krylov projection subspace. Via several examples, we show that the pseudo \mathcal{H}_2 -optimal shifts consistently yield nearly optimal low rank approximations to the solutions of the Lyapunov equations.

1 Introduction

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times m}$ and $\mathbf{Y} \in \mathbb{R}^{n \times m}$ be given matrices. Then, the Sylvester equation for the unknown matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ is given by

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} + \mathbf{Y} = \mathbf{0}. \quad (1)$$

The equation (1) has a unique solution if and only if $\lambda_i(\mathbf{A}) + \lambda_j(\mathbf{B}) \neq 0$ for $i = 1, \dots, n$ and $j = 1, \dots, m$. A special case of the Sylvester equation is the Lyapunov equation, where $\mathbf{B} = \mathbf{A}^*$ and $\mathbf{Y} = \mathbf{Y}^* \geq \mathbf{0}$. Both the Sylvester and Lyapunov equations are an important tool in the analysis of asymptotically stable linear dynamical systems of the form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t), \quad y(t) = \mathbf{c}^*\mathbf{x}(t), \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{b}, \mathbf{c}^* \in \mathbb{R}^n$. In (2), $\mathbf{x}(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, are, respectively the *state*, *input*, and *output*, of the underlying system. While the cross gramian \mathcal{X} of (2) solves the Sylvester equation

$$\mathbf{A}\mathcal{X} + \mathcal{X}\mathbf{A} + \mathbf{b}\mathbf{c}^* = \mathbf{0}, \quad (3)$$

the controllability gramian \mathcal{P} and the observability gramian \mathcal{Q} solve the Lyapunov equations

$$\mathbf{A}\mathcal{P} + \mathcal{P}\mathbf{A}^* + \mathbf{b}\mathbf{b}^* = \mathbf{0} \quad \text{and} \quad \mathbf{A}^*\mathcal{Q} + \mathcal{Q}\mathbf{A} + \mathbf{c}^*\mathbf{c} = \mathbf{0}, \quad (4)$$

respectively. These three gramians are of fundamental importance especially in the concept of model reduction, see [1]. In what follows, we will mainly focus on the Sylvester equation (1) where \mathbf{Y} is rank-1; hence our discussion already contains the Lyapunov equations as a special case.

The standard direct method for solving (1) is due to Bartels and Stewart [3]. However, this method requires dense matrix operations such as the Schur decomposition; thus is not applicable in large-scale settings. For large-scale settings, iterative methods have been developed that take advantage of the sparsity and the low-rank structure of \mathbf{Y} . The two most common ones are the Alternating Direction Implicit (ADI) Method ([30, 8, 9, 27, 17, 37, 33, 21, 31, 43, 39, 44]) and the (rational) Krylov projection methods ([22, 25, 13, 24, 32, 38, 14, 2]).

The ADI method was first introduced by Peaceman and Rachford [29] for solving parabolic and elliptic PDEs, and was later adapted to solving the Sylvester equation by Wachspress in [43]. It is a fixed point iteration scheme for approximating \mathbf{X} . Given two sequences of shifts $\{\alpha_1, \alpha_2, \dots, \alpha_r, \dots\}, \{\beta_1, \beta_2, \dots, \beta_r, \dots\} \subset \mathbb{C}$ and an initial guess \mathbf{X}_0 , the ADI iteration for (1) proceeds as follows :

$$\mathbf{X}_i = (\mathbf{A} - \alpha_i \mathbf{I})(\mathbf{A} + \beta_i \mathbf{I})^{-1} \mathbf{X}_{i-1} (\mathbf{B} - \beta_i \mathbf{I})(\mathbf{B} + \alpha_i \mathbf{I})^{-1} \quad (5)$$

$$- (\alpha_i + \beta_i)(\mathbf{A} + \beta_i \mathbf{I})^{-1} \mathbf{Y} (\mathbf{B} + \alpha_i \mathbf{I})^{-1}. \quad (6)$$

The performance of the ADI iteration depends heavily on the choice of shifts used in the iteration. Several schemes have been developed for making asymptotically optimal shift selections if some information is known about the boundaries of the numerical range of \mathbf{A} , and \mathbf{B} . See [42, 43, 35, 36, 31] and the references therein for further details on the shift selection problem in the ADI iteration.

A closely related method to the ADI iteration is the rational Krylov projection method (RKPM). In the RKPM, the Sylvester equation $\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} + \mathbf{Y} = \mathbf{0}$ is projected onto the rational Krylov subspaces $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \boldsymbol{\sigma}) = \text{span}\{(\sigma_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}, \dots, (\sigma_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\}$ and $\mathcal{K}_r^{\text{rat}}(\mathbf{B}^*, \mathbf{c}, \bar{\boldsymbol{\mu}})$ where $\boldsymbol{\sigma} = \{\sigma_1, \dots, \sigma_r\}$, and $\bar{\boldsymbol{\mu}} = \{\bar{\mu}_1, \dots, \bar{\mu}_r\}$ are the sets of shifts used to construct the respective rational Krylov spaces and $\bar{\nu}$ denotes the conjugate of ν . See [6] for further details regarding $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \boldsymbol{\sigma})$, and constructing an orthonormal basis via the rational Arnoldi iteration. Let \mathbf{Q}_r and \mathbf{U}_r denote the orthonormal basis for $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \boldsymbol{\sigma})$ and $\mathcal{K}_r^{\text{rat}}(\mathbf{B}^*, \mathbf{c}, \bar{\boldsymbol{\mu}})$. Then, the RKPM approximation is constructed by first solving

$$\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r + \tilde{\mathbf{X}}_r \mathbf{U}_r^* \mathbf{B}^* \mathbf{U}_r + \mathbf{Q}_r^* \mathbf{b} \mathbf{c}^* \mathbf{U}_r = \mathbf{0} \quad (7)$$

and then approximating \mathbf{X} by $\mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{U}_r^*$. The solution of the projected Sylvester equation (7) is very cheap. Like the ADI method, the RKPM method also relies heavily on a good choice of shifts to produce accurate results. In the next section we will derive results that show for a certain choice of shifts, the RKPM and ADI methods are indeed equivalent.

Since in almost all applications, the quantities \mathbf{A} , \mathbf{B} , \mathbf{b} , and \mathbf{c} are real, we will assume that the set of shifts $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ are closed under conjugation so that the approximants are real as well. This will guarantee that the orthonormal bases \mathbf{Q}_r for $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \boldsymbol{\sigma})$ and \mathbf{U}_r for $\mathcal{K}_r^{\text{rat}}(\mathbf{B}^*, \mathbf{c}, \bar{\boldsymbol{\mu}})$ can be computed to be real as well.

2 Equivalence of the ADI and Rational Krylov Projection Methods for pseudo- \mathcal{H}_2 optimal points

In this section, we present our main results illustrating the connection between the ADI and RKPM. Since the discussion requires the concept of \mathcal{H}_2 -optimal points for model reduction, we first briefly review the \mathcal{H}_2 approximation problem.

2.1 Optimal \mathcal{H}_2 model reduction

For a full-order model as given in (2), the model reduction problem seeks to construct a dynamical system

$$\dot{\mathbf{x}}_r(t) = \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{b}_r u(t), \quad y_r(t) = \mathbf{c}_r^* \mathbf{x}_r(t) \quad (8)$$

of much smaller dimension $r \ll n$, with $\mathbf{A}_r \in \mathbb{R}^{r \times r}$ and $\mathbf{b}_r, \mathbf{c}_r^* \in \mathbb{R}^r$ such that $y_r(t)$ approximates $y(t)$ well for a wide range of inputs $u(t)$. The reduced-model in (2) is usually obtained via state-space projection: Two matrices $\mathbf{V}_r, \mathbf{W}_r \in \mathbb{R}^{n \times r}$ are constructed with $\mathbf{W}_r^* \mathbf{V}_r = \mathbf{I}_r$ to produce

$$\mathbf{A}_r = \mathbf{W}_r^* \mathbf{A} \mathbf{V}_r, \quad \mathbf{b}_r = \mathbf{W}_r^* \mathbf{b}, \quad \text{and} \quad \mathbf{c}_r = \mathbf{V}_r^* \mathbf{c} \quad (9)$$

One can measure the quality of the approximation using the concept of transfer function. By taking the Laplace transforms of (2) and (8), one obtains the transfer functions $H(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ and $H_r(s) = \mathbf{c}_r(s\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r$, respectively. Hence, one can consider model reduction in terms of these transfer functions as approximating a degree- n rational function $H(s)$ with a degree- r one $H_r(s)$. For more details on model reduction of linear dynamical systems, see [1].

In this paper, we focus on the \mathcal{H}_2 -norm to measure accuracy of the reduced-model. The \mathcal{H}_2 optimal model reduction problem seeks to construct a reduced system as in (8), so that $H_r(s)$ minimizes the \mathcal{H}_2 error over all stable linear dynamical systems of the form (8), i.e.

$$\|H - H_r\|_{\mathcal{H}_2} = \min_{\deg(\tilde{H}_r)=r} \|H - \tilde{H}_r\|_{\mathcal{H}_2} \quad (10)$$

where

$$\|H - H_r\|_{\mathcal{H}_2} = \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega) - H_r(j\omega)|^2 d\omega \right)^{1/2}.$$

Several methods have been introduced to solve (10); see, for example, [34, 23, 45, 46, 28, 19, 18, 41, 11, 16, 4, 5], and the references therein. Since the optimization problem (10) is nonconvex, the common approach involves finding reduced-order models satisfying the first-order necessary conditions of \mathcal{H}_2 -optimality. The next theorem states the interpolation-based necessary conditions for \mathcal{H}_2 optimality introduced by Meier and Luenberger [28].

Theorem 1. [28, 19] *Given a full-order system $H(s)$ of order n , if $H_r(s) = \sum_{i=1}^r \frac{\phi_i}{s - \lambda_i}$ is an \mathcal{H}_2 optimal approximation to $H(s)$, then*

$$H(-\lambda_i) = H_r(-\lambda_i) \quad \text{for} \quad i = 1, \dots, r, \quad \text{and} \quad (11)$$

$$H'(-\lambda_i) = H'_r(-\lambda_i) \quad \text{for} \quad i = 1, \dots, r \quad (12)$$

A reduced-order model which satisfies the \mathcal{H}_2 -optimality conditions can be obtained by using the Iterative Rational Krylov Algorithm (IRKA) of Gugercin et. al. in [19]. However, in this paper we will focus on satisfying only (11) (without the derivative condition). We will call these interpolation points *pseudo \mathcal{H}_2 -optimal points* to emphasize that they only satisfy a subset of the optimality conditions. In terms of the projection framework (9) for model reduction, this corresponds to finding interpolation points $\sigma = \{\sigma_1, \dots, \sigma_r\}$ and choosing, in (9), $V_r = W_r = Q_r$ where Q_r is an orthonormal basis for the rational Krylov subspace $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \sigma)$ so that $\{\lambda_1, \dots, \lambda_r\}$, i.e. the eigenvalues of $\mathbf{A}_r = Q_r^T \mathbf{A} Q_r$, become the mirror images of the interpolations points $\sigma = \{\sigma_1, \dots, \sigma_r\}$, i.e.

$$\lambda(\mathbf{A}_r) = \lambda(Q_r^T \mathbf{A} Q_r) = -\sigma. \quad (13)$$

The emhpseudo \mathcal{H}_2 -optimal points interpolation points can be computed iteratively in a manner similar to IRKA [19] as done in [20] for port-Hamiltonian systems.

2.2 The ADI Iteration and Rational Krylov Projection Method

The main theorem requires the following lemma, which connects the ADI approximation for the Sylvester equation with rational Krylov subspaces. This extends an earlier result by Li and White [26] which establishes a similar connection for the the case of the Lyapunov equation.

Lemma 1. *Let $\mathbf{Y} = \mathbf{b}\mathbf{c}^*$, where $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{c} \in \mathbb{R}^m$. Let $\{\sigma_1, \dots, \sigma_r\}$ and $\{\mu_1, \dots, \mu_r\}$ be two collections of shifts that satisfy $\Re(\mu_i), \Re(\sigma_i) > 0$ for $i = 1, \dots, r$. Suppose \mathbf{X}_r is the approximate solution to the Sylvester equation (1) obtained by applying the pair of shifts $\alpha_i = -\sigma_i$ and $\beta_i = -\mu_i$ in the ADI iteration (5) for $i = 1, \dots, r$ with $\mathbf{X}_0 = \mathbf{0}$. Then there exist $\mathbf{L}_r \in \mathbb{C}^{n \times r}$ and $\mathbf{R}_r \in \mathbb{C}^{m \times r}$ such that $\mathbf{X}_r = \mathbf{L}_r \mathbf{R}_r^*$ and $\text{colspan}(\mathbf{L}_r) \subset \mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \mu)$ and $\text{colspan}(\mathbf{R}_r) \subset \mathcal{K}_r^{\text{rat}}(\mathbf{B}^*, \mathbf{c}, \bar{\sigma})$*

Proof. The proof is given by induction on i , the iteration step. First note that for $i = 1$, $\mathbf{X}_1 = (\mu_1 + \sigma_1)(\mathbf{A} - \mu_1 \mathbf{I})^{-1} \mathbf{b}\mathbf{c}^*(\mathbf{B} - \sigma_1 \mathbf{I})^{-1}$, so let $\mathbf{L}_1 = [(\mu_1 + \sigma_1)(\mathbf{A} - \mu_1 \mathbf{I})^{-1} \mathbf{b}]$ and $\mathbf{R}_1 = [(\mathbf{B}^* - \bar{\sigma}_1 \mathbf{I})^{-1} \mathbf{c}]$. Then \mathbf{L}_1 and \mathbf{R}_1 clearly satisfy the hypothesis and $\mathbf{X}_1 = \mathbf{L}_1 \mathbf{R}_1^*$. Now suppose that the statement holds for \mathbf{X}_i . Then, for $j = 1, \dots, i$, the j^{th} column of \mathbf{L}_i is $T_i^{(j)}(\mathbf{A})\mathbf{b}$, where $T_i^{(j)}(\lambda)$ is a proper rational function that lies in the span of $\{\frac{1}{\lambda - \mu_1}, \dots, \frac{1}{\lambda - \mu_i}\}$. Similarly, the j^{th} column of \mathbf{R}_i is $S_i^{(j)}(\mathbf{B}^*)\mathbf{c}$, where $S_i^{(j)}(\lambda)$ lies in the span of $\{\frac{1}{\lambda - \bar{\sigma}_1}, \dots, \frac{1}{\lambda - \bar{\sigma}_i}\}$. Therefore \mathbf{X}_{i+1} can be written as

$$\begin{aligned} \mathbf{X}_{i+1} &= (\mathbf{A} + \sigma_{i+1} \mathbf{I})(\mathbf{A} - \mu_{i+1} \mathbf{I})^{-1} \mathbf{L} \mathbf{R}^* (\mathbf{B} + \mu_{i+1} \mathbf{I})(\mathbf{B} - \sigma_{i+1} \mathbf{I})^{-1} \\ &\quad + (\mu_{i+1} + \sigma_{i+1})(\mathbf{A} - \mu_{i+1} \mathbf{I})^{-1} \mathbf{b}\mathbf{c}^*(\mathbf{B} - \sigma_{i+1} \mathbf{I})^{-1} \\ &= \sum_{j=1}^i (\mathbf{A} + \sigma_{i+1} \mathbf{I})(\mathbf{A} - \mu_{i+1} \mathbf{I})^{-1} T_i^{(j)}(\mathbf{A}) \mathbf{b}\mathbf{c}^* S_i^{(j)}(\mathbf{B}) (\mathbf{B} + \mu_{i+1} \mathbf{I})(\mathbf{B} - \sigma_{i+1} \mathbf{I})^{-1} \\ &\quad + (\mu_{i+1} + \sigma_{i+1})(\mathbf{A} - \mu_{i+1} \mathbf{I})^{-1} \mathbf{b}\mathbf{c}^*(\mathbf{B} - \sigma_{i+1} \mathbf{I})^{-1} \end{aligned}$$

For $j = 1, \dots, i$, let the j^{th} column of \mathbf{L}_{i+1} be $(\mathbf{A} + \sigma_{i+1} \mathbf{I})(\mathbf{A} - \mu_{i+1} \mathbf{I})^{-1} T_i^{(j)}(\mathbf{A})\mathbf{b}$ and let the $(i+1)^{\text{th}}$ column be $T_{i+1}^{(i+1)}(\mathbf{A})\mathbf{b} = (\mu_{i+1} + \sigma_{i+1})(\mathbf{A} - \mu_{i+1} \mathbf{I})^{-1} \mathbf{b}$. Then clearly $\text{colspan}(\mathbf{L}_{i+1}) \subset \mathcal{K}_{i+1}^{\text{rat}}(\mathbf{A}, \mathbf{b}, \mu)$. Similarly, let $(\mathbf{B}^* - \bar{\sigma}_{i+1} \mathbf{I})^{-1} (\mathbf{B}^* + \bar{\mu}_{i+1} \mathbf{I}) S_i^{(j)}(\mathbf{B}^*)\mathbf{c}$ be the j^{th} column of \mathbf{R}_{i+1} for $j = 1, \dots, i$, and $S_{i+1}^{(i+1)}(\mathbf{B}^*)\mathbf{c} = (\mathbf{B}^* - \bar{\sigma}_{i+1} \mathbf{I})^{-1} \mathbf{c}$ be the $(i+1)^{\text{th}}$ column. Then

$\text{colspan}(\mathbf{R}_{i+1}) \subset \mathcal{K}_{i+1}^{\text{rat}}(\mathbf{B}^*, \mathbf{c}, \bar{\sigma})$. Finally, we note that by construction, $\mathbf{X}_{i+1} = \mathbf{L}_{i+1} \mathbf{R}_{i+1}^*$. \square

Next, we give our first main result showing that the approximate solution of the Sylvester equation (1) by ADI and RKPM are indeed equivalent when the shifts are chosen as pseudo- \mathcal{H}_2 optimal points. This result applied to the special case of Lyapunov equation was first presented at the 2010 SIAM Annual Meeting [15] then later published independently in [13]. Our new result here, on the other hand, is more general than both [15] and [13] since it tackles the case of Sylvester equation and includes the Lyapunov equation as a special case. Moreover, while the proof given in [13] for the special case of Lyapunov equation makes use of a novel connection between the ADI iteration and the so-called Skeleton approximation framework first developed in the work of Tyrtyshnikov [40], the proof we provide here for the more general Sylvester equation case is given directly in terms of rational Krylov interpolation conditions, and in that sense is simpler.

Theorem 2. *Given the Sylvester equation (1) with $\mathbf{Y} = \mathbf{b}\mathbf{c}^*$, where $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{c} \in \mathbb{R}^m$, let $\mathbf{Q}_r \in \mathbb{R}^{n \times r}$ be an orthonormal basis for the rational Krylov subspace $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \sigma)$ and let $\mathbf{U}_r \in \mathbb{R}^{m \times r}$ be an orthonormal basis for the rational Krylov subspace $\mathcal{K}_r^{\text{rat}}(\mathbf{B}^*, \mathbf{c}, \bar{\sigma})$ for a set of shifts $\sigma = \{\sigma_1, \dots, \sigma_r\}$ where $\Re(\sigma_i) > 0$ for $i = 1, \dots, r$. Let $\tilde{\mathbf{X}}_r \in \mathbb{R}^{r \times r}$ solve the projected Sylvester equation*

$$\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r + \tilde{\mathbf{X}}_r \mathbf{U}_r^* \mathbf{B} \mathbf{U}_r + \mathbf{Q}_r^* \mathbf{b} \mathbf{c}^* \mathbf{U}_r = \mathbf{0}, \quad (14)$$

and let $\mathbf{X}_r \in \mathbb{R}^{n \times m}$ be computed by applying the shifts $\alpha_i = -\sigma_i$ and $\beta_i = -\sigma_i$ to exactly r steps of the ADI iteration (5) for $i = 1, \dots, r$. Then $\mathbf{X}_r = \mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{U}_r^*$ if and only if either $\lambda(\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r) = -\{\sigma_1, \dots, \sigma_r\}$ or $\lambda(\mathbf{U}_r^* \mathbf{B} \mathbf{U}_r) = -\{\sigma_1, \dots, \sigma_r\}$.

Proof. (\Leftarrow) First suppose that $\lambda(\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r) = -\{\sigma_1, \dots, \sigma_r\}$. The proof remains the same if we instead suppose that $\lambda(\mathbf{U}_r^* \mathbf{B} \mathbf{U}_r) = -\{\sigma_1, \dots, \sigma_r\}$. Let $\mathbf{A}_r = \mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r$, and $\mathbf{b}_r = \mathbf{Q}_r^* \mathbf{b}$, $\mathbf{B}_r = \mathbf{U}_r^* \mathbf{B} \mathbf{U}_r$, and $\mathbf{c}_r = \mathbf{U}_r^* \mathbf{c}$. Note that after we apply r steps of the ADI iteration with the set of shifts $\alpha_i = \beta_i = -\sigma_i$ to the projected Sylvester equation (14), we obtain the exact solution $\tilde{\mathbf{X}}_r$, since $\lambda(\mathbf{A}_r) = -\{\sigma_1, \dots, \sigma_r\}$. By Lemma 1, at the r th step of the ADI iteration $\tilde{\mathbf{X}}_r = \tilde{\mathbf{L}}_r \tilde{\mathbf{R}}_r^*$ where $\tilde{\mathbf{L}}_r = [T^{(1)}(\mathbf{A}_r) \mathbf{b}_r, \dots, T^{(r)}(\mathbf{A}_r) \mathbf{b}_r]$ where $T^{(i)}(\mathbf{A}_r) \mathbf{b}_r$ are rational functions that lie in $\mathcal{K}_r^{\text{rat}}(\mathbf{A}_r, \mathbf{b}_r, \sigma)$. Similarly $\tilde{\mathbf{R}}_r = [S^{(1)}(\mathbf{B}_r^*) \mathbf{c}_r^*, \dots, S^{(r)}(\mathbf{B}_r^*) \mathbf{c}_r^*]$ where the $S^{(i)}(\mathbf{B}_r^*) \mathbf{c}_r^*$ are rational functions that lie in $\mathcal{K}_r^{\text{rat}}(\mathbf{B}_r^*, \mathbf{c}_r^*, \bar{\sigma})$. Furthermore, for the same shifts, $\alpha_i = \beta_i = -\sigma_i$ for $i = 1, \dots, r$, applied to r steps of the ADI iteration on the full Sylvester equation (1), we have $\mathbf{X}_r = \mathbf{L}_r \mathbf{R}_r^*$ and $\mathbf{L}_r = [T^{(1)}(\mathbf{A}) \mathbf{b}, \dots, T^{(r)}(\mathbf{A}) \mathbf{b}]$ and $\mathbf{R}_r = [S^{(1)}(\mathbf{B}^*) \mathbf{c}, \dots, S^{(r)}(\mathbf{B}^*) \mathbf{c}]$. Thus it is sufficient to show that $\mathbf{Q}_r \tilde{\mathbf{L}}_r = \mathbf{L}_r$ and that $\mathbf{U}_r \tilde{\mathbf{R}}_r = \mathbf{R}_r$. Without loss of generality consider just the former equation. This, in turn, amounts to showing that $\mathbf{Q}_r T^{(i)}(\mathbf{A}_r) \mathbf{b}_r = T^{(i)}(\mathbf{A}) \mathbf{b}$. If $T_i(\mathbf{A}) \mathbf{b}$ are a set of orthogonal rational functions that span $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \sigma)$, then it is sufficient to show that

$$\mathbf{Q}_r T_i(\mathbf{A}_r) \mathbf{b}_r = T_i(\mathbf{A}) \mathbf{b}. \quad (15)$$

Equality (15) follows readily from the interpolation properties of the Galerkin projection, which we show below. First, note that due to the interpolation properties of the Galerkin projection, $\mathbf{Q}_r (\sigma_i \mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r = (\sigma_i \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}$. Let $\mathbf{V}_r = [(\sigma_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \dots (\sigma_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}]$. Then, for some $\mathbf{x} \in \mathbb{R}^r$,

$$\mathbf{V}_r \mathbf{x} = T_i(\mathbf{A}) \mathbf{b} = \mathbf{Q}_r [(\sigma_1 \mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r \dots (\sigma_r \mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r] \mathbf{x} = \mathbf{Q}_r T_i(\mathbf{A}_r) \mathbf{b}_r, \quad (16)$$

which proves (15).

(\Rightarrow) Let $\tilde{\mathbf{X}}_r$ be the solution of

$$\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r + \tilde{\mathbf{X}}_r \mathbf{U}_r^* \mathbf{B} \mathbf{U}_r + \mathbf{Q}_r^* \mathbf{b} \mathbf{c}^* \mathbf{U}_r = \mathbf{0} \quad (17)$$

where \mathbf{Q}_r is an orthonormal basis for $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \boldsymbol{\sigma})$ and \mathbf{U}_r is an orthonormal basis for $\mathcal{K}_r^{\text{rat}}(\mathbf{B}^*, \mathbf{c}, \bar{\boldsymbol{\sigma}})$. Suppose that $\mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{U}_r^* = \mathbf{X}_r$. Let $\hat{\mathbf{X}}_r$ be the approximate solution of (17) resulting from applying the shifts $\alpha_i = \beta_i = -\sigma_i$ for $i = 1, \dots, r$ to exactly r steps of the ADI iteration (5). By the interpolation result given in the proof above, $\mathbf{Q}_r \hat{\mathbf{X}}_r \mathbf{U}_r^* = \mathbf{X}_r$. It follows from the assumptions that, $\mathbf{Q}_r \hat{\mathbf{X}}_r \mathbf{U}_r^* = \mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{U}_r^*$, so $\hat{\mathbf{X}}_r = \tilde{\mathbf{X}}_r$. But this means that $\tilde{\mathbf{X}}_r$ solves (17), and so either $\lambda(\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r) = -\{\sigma_1, \dots, \sigma_r\}$ or $\lambda(\mathbf{U}_r^* \mathbf{B} \mathbf{U}_r) = -\{\sigma_1, \dots, \sigma_r\}$. \square

Remark 1. This theorem shows that the ADI approximation for the Sylvester equation is equivalent to lifting the solution of the projected Sylvester equation back to the original dimension when either $\lambda(\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r) = -\{\sigma_1, \dots, \sigma_r\}$ or $\lambda(\mathbf{U}_r^* \mathbf{B} \mathbf{U}_r) = -\{\sigma_1, \dots, \sigma_r\}$; hence the two most common approximation methods for solving a Sylvester equation is indeed equivalent for these special shift selection. Recalling pseudo- \mathcal{H}_2 optimality condition (13), for a given r , these special shifts are indeed exactly the pseudo- \mathcal{H}_2 optimal shifts for a dynamical system $H_1(s) = \mathbf{z}_1(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ or $H_2(s) = \mathbf{z}_2(s\mathbf{I} - \mathbf{B}^*)^{-1}\mathbf{c}^*$ where \mathbf{z}_1 and \mathbf{z}_2 are vectors of appropriate sizes.

2.3 Orthogonality in the case of Lyapunov equation

The parameters for which the ADI iteration and the rational Krylov projections coincide also satisfy orthogonality conditions on the residual for the special case of the Lyapunov equation

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^* + \mathbf{b}\mathbf{b}^* = \mathbf{0} \quad (18)$$

For a given approximation \mathbf{X}_r to the solution \mathbf{X} , define the residual \mathbf{R} as

$$\mathbf{R} = \mathbf{A}\mathbf{X}_r + \mathbf{X}_r\mathbf{A}^* + \mathbf{b}\mathbf{b}^*. \quad (19)$$

The following result was first given in [13]. Here we present a new and more concise proof of the orthogonality result in terms of the special interpolation properties of the pseudo \mathcal{H}_2 -optimal shifts.

Theorem 3. Given $\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^* + \mathbf{b}\mathbf{b}^* = \mathbf{0}$, let $\tilde{\mathbf{X}}_r \in \mathbb{R}^{r \times r}$ solve the projected Lyapunov equation

$$\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r + \tilde{\mathbf{X}}_r \mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r + \mathbf{Q}_r^* \mathbf{b} \mathbf{b}^* \mathbf{Q}_r = \mathbf{0},$$

where \mathbf{Q}_r is an orthonormal basis for the $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \boldsymbol{\sigma})$ with $\boldsymbol{\sigma} = \{\sigma_1, \dots, \sigma_r\}$. Let $\mathbf{X}_r = \mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{Q}_r^*$. Then $\mathbf{Q}_r^* \mathbf{R} = \mathbf{0}$ if and only if $\lambda(\mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r) = -\{\sigma_1, \dots, \sigma_r\}$ where \mathbf{R} is the residual defined in (19).

Proof. (\Rightarrow) Suppose that $\mathbf{Q}_r^* \mathbf{R} = \mathbf{0}$. Multiplying (19) with \mathbf{Q}_r^* from the left and then transposing the resulting equation leads to

$$\mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r + \mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{Q}_r^* \mathbf{A}^* \mathbf{Q}_r + \mathbf{b} \mathbf{b}^* \mathbf{Q}_r = \mathbf{0}. \quad (20)$$

Let $\mathbf{A}_r = \mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r = \mathbf{T} \boldsymbol{\Lambda} \mathbf{T}^{-1}$ be the eigenvalue decomposition of \mathbf{A}_r where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_r)$. Plug these expressions into (20), and right multiply by \mathbf{T}^{-*} to obtain

$$\mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{T}^{-*} \boldsymbol{\Lambda}^* + \mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{T}^{-*} + \mathbf{b} \mathbf{b}^* \mathbf{Q}_r \mathbf{T}^{-*} = \mathbf{0} \quad (21)$$

Let ζ_i be the i^{th} entry of $\mathbf{b}^* \mathbf{Q}_r \mathbf{T}^{-*}$. Then it is straightforward to show that the i^{th} column of $\mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{T}^{-*}$ must be $(-\bar{\lambda}_i \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \zeta_i$. Thus, it follows that $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, \boldsymbol{\sigma}) = \mathcal{K}_r^{\text{rat}}(\mathbf{A}, \mathbf{b}, -\bar{\boldsymbol{\lambda}})$, where $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_r\}$. Since both sets $\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ are closed under conjugation, after an appropriate reordering, we obtain $\sigma_i = -\lambda_i$.

(\Leftarrow) Observe that

$$\mathbf{A}_r \tilde{\mathbf{X}}_r + \tilde{\mathbf{X}}_r \mathbf{A}_r^* + \mathbf{Q}_r^* \mathbf{b} \mathbf{b}^* \mathbf{Q}_r = \mathbf{0} \Rightarrow \quad (22)$$

$$\mathbf{A}_r \tilde{\mathbf{X}}_r \mathbf{T}^{-*} + \tilde{\mathbf{X}}_r \mathbf{T}^{-*} \boldsymbol{\Lambda}^* + \mathbf{Q}_r^* \mathbf{b} \mathbf{b}^* \mathbf{Q}_r \mathbf{T}^{-*} = \mathbf{0}. \quad (23)$$

Thus, the i th column of $\tilde{\mathbf{X}}_r \mathbf{T}^{-*}$ is $(-\bar{\lambda}_i \mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{Q}_r^* \mathbf{b} \zeta_i$. But since \mathbf{Q}_r is an orthonormal basis for $\mathcal{K}_r^{\text{rat}}(\mathbf{A}, \tilde{\mathbf{b}}, \boldsymbol{\sigma})$, and $\lambda_i = -\bar{\sigma}_i$, this means

$$\mathbf{Q}_r (-\bar{\lambda}_i \mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{Q}_r^* \mathbf{b} \zeta_i = (-\bar{\lambda}_i \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \zeta_i = (\mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{T}^{-*}) \mathbf{e}_i, \quad (24)$$

where \mathbf{e}_i is the i th unit vector. Thus,

$$\mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{T}^{-*} \boldsymbol{\Lambda}^* + \mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{T}^{-*} + \mathbf{b} \mathbf{b}^* \mathbf{Q}_r \mathbf{T}^{-*} = \mathbf{0}, \quad (25)$$

which implies

$$\mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{Q}_r^* \mathbf{A}^* \mathbf{Q}_r + \mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r + \mathbf{b} \mathbf{b}^* \mathbf{Q}_r = \mathbf{0}. \quad (26)$$

Transpose this last expression and use the fact that $\mathbf{Q}_r^* \mathbf{Q}_r = \mathbf{I}_r$ to obtain

$$\mathbf{Q}_r^* \mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{Q}_r^* \mathbf{A}^* + \mathbf{Q}_r^* \mathbf{A} \mathbf{Q}_r \tilde{\mathbf{X}}_r \mathbf{Q}_r^* + \mathbf{Q}_r^* \mathbf{b} \mathbf{b}^* = \mathbf{Q}_r^* \mathbf{R} = \mathbf{0}, \quad (27)$$

which is the desired result. \square

Remark 2. As we have previously noted, in almost every practical situation, one would choose a set of shifts $\boldsymbol{\sigma}$ which is closed under conjugation. But even for the cases where this assumption on $\boldsymbol{\sigma}$ does not hold, Theorem 2 holds as is, and Theorem 3 applies with a slight modification. To wit, $\lambda_i = \bar{\sigma}_i$ for $i = 1, \dots, r$.

3 A numerical study on using the pseudo- \mathcal{H}_2 optimal points as the ADI shifts

Having shown that using the pseudo- \mathcal{H}_2 optimal points in the ADI iteration for the Sylvester equation is equivalent to applying RKPM and that the pseudo- \mathcal{H}_2 points leads to an orthogonality condition in the case of Lyapunov equation, the natural question to ask is what quality of approximation the pseudo- \mathcal{H}_2 optimal points have as ADI shifts. We will briefly investigate this issue in this section. However, we emphasize that the purpose of our numerical results is not to advocate employing the pseudo \mathcal{H}_2 -optimal shifts in the ADI iteration or in the RKPM. This would be a costly numerical method for approximating Sylvester equations since obtaining the pseudo \mathcal{H}_2 -optimal shifts already requires solving several linear systems. Our numerical results are meant to illustrate the unique quality of these shifts compared with other choices of shifts that do not share the ADI-RKPM equivalency property.

We used three benchmark models in our numerical simulations: The CD Player model with $n = 120$, the EADY model with $n = 598$, and the Rail Model with $n = 1357$. The first two models are described in detail in [12] and the Rail model in [7]. For all three models,

we compute a rank r approximation to the solution of the the Lyapunov equation (18). The exact and approximate solutions are denoted by \mathbf{X} and \mathbf{X}_r , respectively. The Rail model has multiple inputs; thus for this model we only use the first column of the input matrix. We use three different approximation methods for each model:

- Method 1: The RKPM is applied to the a sequence of shifts that alternates between 0 and ∞ . The resulting subspace is generally referred to as the extended Krylov subspace. Its application to RKPM was first introduced by Simoncini in [32].
- Method 2: The RKPM is applied using r pseudo- \mathcal{H}_2 optimal shifts; or equivalently r steps the ADI iteration is applied using r pseudo- \mathcal{H}_2 optimal shifts.
- Method 3: The r -steps of ADI iteration are applied where the ADI shifts are chosen via Penzl’s heuristic method [30].

The quality of the resulting approximations from each method are compared using the relative error in the 2-norm, i.e. $\frac{\|\mathbf{X} - \mathbf{X}_r\|_2}{\|\mathbf{X}\|_2}$. Figure 1 shows the relative errors for the EADY model as r varies from 1 to 50 together with the minimum possible error, i.e. $\frac{\pi_{r+1}}{\pi_1}$ where π_i is the i^{th} singular value of the true solution \mathbf{X} . Note that for a given r , the pseudo- \mathcal{H}_2 optimal shifts perform remarkably well, almost matching the best low-rank approximation given by the singular value decomposition. For a selected number of r values, these numbers are also tabulated in Table 1 further illustrating the effectiveness of the pseudo- \mathcal{H}_2 points as ADI or RKPM shifts. Similar results for the CD player model are shown in Figure 2 and in Table 2 and for the Rail Model in Figure 3 and Table 3 illustrating that the pseudo \mathcal{H}_2 -optimal shifts produce a nearly optimal rank r approximation in several cases. Indeed, this phenomenon was recently explained by Breiten and Benner in [10], where they show that the \mathcal{H}_2 optimal shifts are optimal in a special energy norm related to the Lyapunov equation; for details we refer the reader to [10].

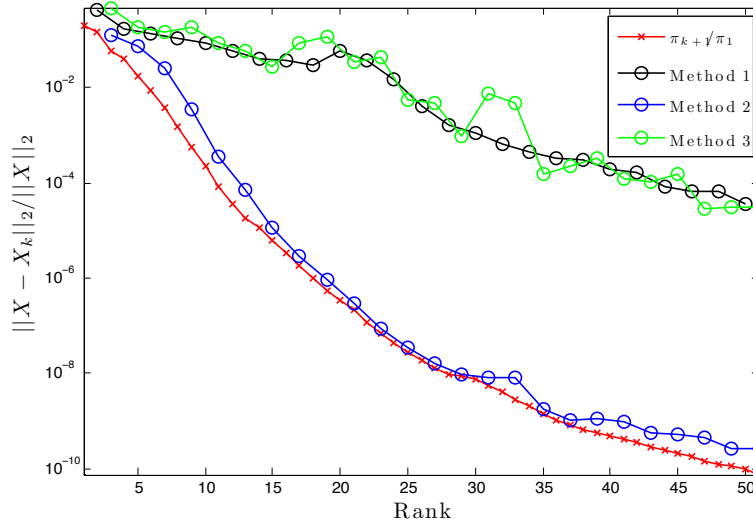


Figure 1: Relative error as r varies for the EADY Model

Table 1: Comparison of Methods with the EADY model order 598

r	$\frac{\pi_{r+1}}{\pi_1}$	$\frac{\ \mathbf{X} - \mathbf{X}_r\ _2}{\ \mathbf{X}\ _2}$		
		Method 1	Method 2	Method 3
10	2.31×10^{-4}	8.42×10^{-2}	1.28×10^{-3}	1.96×10^{-1}
20	3.38×10^{-7}	5.73×10^{-2}	4.99×10^{-7}	1.13×10^{-1}
30	7.63×10^{-9}	1.09×10^{-3}	8.46×10^{-9}	8.54×10^{-2}
40	4.83×10^{-10}	1.93×10^{-4}	1.06×10^{-9}	9.70×10^{-3}

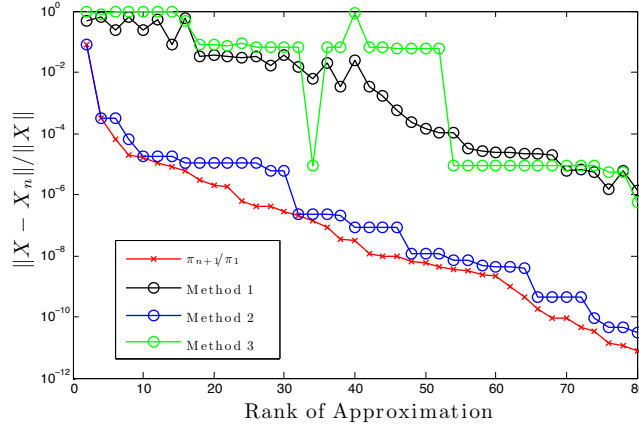


Figure 2: Relative error as r varies for for the CD Player model

4 Acknowledgements

The authors thank Prof. Christopher Beattie of Virginia Tech. for several fruitful discussions on this subject. This work has been supported in part by the NSF Grant DMS-0645347.

5 Conclusions

In this paper we presented a new result that solidifies the connection between the ADI iteration and rational Krylov projection methods for solving large-scale Sylvester equation. We have shown that for one-sided projections, the two methods are indeed equivalent for a special choice of shifts called pseudo- \mathcal{H}_2 optimal shifts, so-called because they partially satisfy first-order necessary conditions for \mathcal{H}_2 optimal model reduction. These shifts are also optimal in the sense that they produce an approximation with a residual orthogonal to the rational Krylov projection subspace in the case of Lyapunov equation.

References

- [1] A.C. Antoulas. *Approximation of Large-Scale Dynamical Systems (Advances in Design and Control)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.

Table 2: Comparison of Errors for the CD Player model

r	$\frac{\pi_{r+1}}{\pi_1}$	$\frac{\ \mathbf{X} - \mathbf{X}_r\ _2}{\ \mathbf{X}\ _2}$		
		Method 1	Method 2	Method 3
4	3.20×10^{-4}	6.71×10^{-1}	3.20×10^{-4}	8.04×10^{-1}
10	1.68×10^{-5}	2.39×10^{-1}	1.73×10^{-5}	9.92×10^{-1}
20	1.92×10^{-6}	3.70×10^{-2}	1.10×10^{-5}	7.84×10^{-2}
40	3.22×10^{-8}	1.62×10^{-3}	8.24×10^{-8}	8.56×10^{-2}

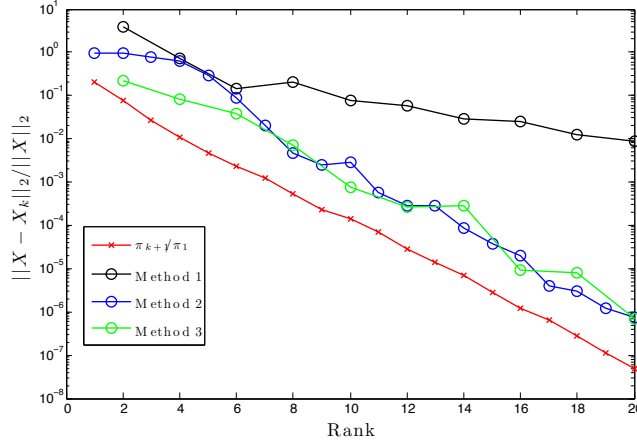


Figure 3: Relative error as r varies for for the Rail Model

- [2] L. Bao, Y. Lin, and Y. Wei. A new projection method for solving large Sylvester equations. *Applied numerical mathematics*, 57(5-7):521–532, 2007.
- [3] R.H. Bartels and GW Stewart. Algorithm 432: Solution of the matrix equation $AX + XB = C$. *Communications of the ACM*, 15(9):820–826, 1972.
- [4] C.A. Beattie and S. Gugercin. Krylov-based minimization for optimal \mathcal{H}_2 model reduction. *46th IEEE Conference on Decision and Control*, pages 4385–4390, Dec. 2007.
- [5] C.A. Beattie and S. Gugercin. A trust region method for optimal \mathcal{H}_2 model reduction. *48th IEEE Conference on Decision and Control*, Dec. 2009.
- [6] B. Beckermann, S. Güttel, and R. Vandebril. On the convergence of rational Ritz values. *SIAM Journal on Matrix Analysis and Applications*, 31(4):1740–1774, 2010.
- [7] P. Benner and J. Saak. Efficient numerical solution of the LQR-problem for the heat equation. *Proc. Appl. Math. Mech*, 4(1):648–649, 2004.
- [8] P. Benner, E. S. Quintana-Ortí, and G. Quintana-Ortí. State-Space Truncation Methods for Parallel Model Reduction of Large-Scale Systems. *Parallel Computing, special issue on “Parallel and Distributed Scientific and Engineering Computing”*, 29:1701–1722, 2003.

Table 3: Comparison of Errors for the Rail Model

r	$\frac{\pi_{r+1}}{\pi_1}$	$\frac{\ \mathbf{X} - \mathbf{X}_r\ _2}{\ \mathbf{X}\ _2}$		
		Method 1	Method 2	Method 3
2	7.47×10^{-2}	3.83×10^0	9.47×10^{-1}	2.14×10^{-1}
6	2.31×10^{-3}	1.44×10^{-1}	8.86×10^{-2}	3.80×10^{-2}
14	6.76×10^{-6}	2.88×10^{-2}	8.43×10^{-5}	2.90×10^{-4}
20	4.96×10^{-8}	8.89×10^{-3}	7.29×10^{-7}	6.85×10^{-7}

- [9] P. Benner, R.-C. Li., and N Truhar. On the ADI method for Sylvester equations. *Journal of Computational and Applied Mathematics*, 233(4):1035 – 1045, 2009.
- [10] Peter Benner and Tobias Breiten. On optimality of interpolation-based low-rank approximations of large-scale matrix equations. *Max Planck Institute Magdeburg Preprints*, December, 2011.
- [11] A. Bunse-Gerstner, D. Kubalinska, G. Vossen, and D. Wilczek. \mathcal{H}_2 -norm optimal model reduction for large scale discrete dynamical MIMO systems. *Journal of computational and applied mathematics*, 233(5):1202–1216, 2010.
- [12] Y. Chahlaoui and P. Van Dooren. Benchmark examples for model reduction of linear time-invariant dynamical systems. *Dimension Reduction of Large-Scale Systems*, 45: 381–395, 2005.
- [13] V. Druskin, L. Knizhnerman, and V. Simoncini. Analysis of the Rational Krylov Subspace and ADI Methods for Solving the Lyapunov Equation. *SIAM Journal on Numerical Analysis*, 49(5):1875–1898, 2011.
- [14] A. El Guennouni, K. Jbilou, and AJ Riquet. Block Krylov subspace methods for solving large Sylvester equations. *Numerical Algorithms*, 29(1):75–96, 2002.
- [15] G. M. Flagg. \mathcal{H}_2 -optimal interpolation: New properties and applications, July, 2010. Talk given at the 2010 SIAM Annual Meeting, Pittsburgh (PA).
- [16] S. Gugercin. An iterative rational Krylov algorithm (IRKA) for optimal \mathcal{H}_2 model reduction. In *Householder Symposium XVI*, Seven Springs Mountain Resort, PA, USA, May 2005.
- [17] S. Gugercin, D.C. Sorensen, and A.C. Antoulas. A modified low-rank Smith method for large-scale Lyapunov equations. *Numerical Algorithms*, 32(1):27–55, 2003.
- [18] S. Gugercin, A. Antoulas, and C. Beattie. A rational Krylov iteration for optimal \mathcal{H}_2 model reduction. In *Proceedings of MTNS*, volume 2006, 2006.
- [19] S. Gugercin, A.C. Antoulas, and C. Beattie. \mathcal{H}_2 model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications*, 30(2):609–638, 2008.

- [20] S. Gugercin, R.V. Polyuga, C.A. Beattie, and A. van der Schaft. Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems. *Automatica*, 2011. Accepted to appear. Available as arXiv:1101.3485v2.
- [21] M. Heinkenschloss, D.C. Sorensen, and K. Sun. Balanced Truncation Model Reduction for a Class of Descriptor Systems with Application to the Oseen Equations. *SIAM Journal on Scientific Computing*, 30:1038, 2008.
- [22] D.Y. Hu and L. Reichel. Krylov-subspace methods for the Sylvester equation. *Linear Algebra and its Applications*, 172:283–313, 1992.
- [23] D. Hyland and D. Bernstein. The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore. *IEEE Trans. Automatic Control*, 30(12):1201–1211, 1985.
- [24] I.M. Jaimoukha and E.M. Kasenally. Krylov subspace methods for solving large Lyapunov equations. *SIAM Journal on Numerical Analysis*, pages 227–251, 1994.
- [25] K. Jbilou. Low rank approximate solutions to large Sylvester matrix equations. *Applied mathematics and computation*, 177(1):365–376, 2006.
- [26] J.R. Li and J. White. Low Rank Solution of Lyapunov Equations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):260–280, 2002.
- [27] J.R. Li and J. White. Low-rank solution of Lyapunov equations. *SIAM review*, pages 693–713, 2004.
- [28] L. Meier III and D. Luenberger. Approximation of linear constant systems. *Automatic Control, IEEE Transactions on*, 12(5):585–588, 1967.
- [29] D.W. Peaceman and HH Rachford. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, 3(1):28–41, 1955.
- [30] T. Penzl. A cyclic low rank Smith method for large sparse Lyapunov equations. *SIAM Journal on Scientific Comput*, 21(4):1401–1418, 2000.
- [31] J. Sabino. *Solution of large-scale Lyapunov equations via the block modified Smith method*. PhD thesis, RICE UNIVERSITY, 2007.
- [32] V. Simoncini. A new iterative method for solving large-scale Lyapunov matrix equations. *SIAM Journal on Scientific Computing*, 29(3):1268–1288, 2008.
- [33] D.C. Sorensen and A.C. Antoulas. The Sylvester equation and approximate balanced reduction. *Linear algebra and its applications*, 351:671–700, 2002.
- [34] J.T. Spanos, M.H. Milman, and D.L. Mingori. A new algorithm for L^2 optimal model reduction. *Automatica*, 28(5):897–909, 1992.
- [35] G. Starke. Optimal alternating direction implicit parameters for nonsymmetric systems of linear equations. *SIAM journal on numerical analysis*, pages 1431–1445, 1991.

- [36] G. Starke. Fejer-Walsh points for rational functions and their use in the ADI iterative method. *Journal of Computational and Applied Mathematics*, 46(1-2):129–141, 1993.
- [37] T. Stykel. Gramian-Based Model Reduction for Descriptor Systems. *Mathematics of Control, Signals, and Systems (MCSS)*, 16(4):297–319, 2004.
- [38] T. Stykel and V. Simoncini. Krylov subspace methods for projected Lyapunov equations. *Applied Numerical Mathematics*, 2011.
- [39] N. Truhar and R.C. Li. On the ADI Method for Sylvester Equations. Technical report, Technical Report 2008-02, Department of Mathematics, University of Texas at Arlington, 2008, available at http://www.uta.edu/math/preprint/rep2008_02.pdf, 2007.
- [40] E. Tyrtshnikov. Mosaic-skeleton approximations. *Calcolo*, 33(1):47–57, 1996.
- [41] P. Van Dooren, K.A. Gallivan, and P.A. Absil. \mathcal{H}_2 -optimal model reduction of MIMO systems. *Applied Mathematics Letters*, 21(12):1267–1273, 2008.
- [42] E.L. Wachspress. Extended application of alternating direction implicit iteration model problem theory. *Journal of the Society for Industrial and Applied Mathematics*, 11(4):994–1016, 1963.
- [43] E.L. Wachspress. The ADI minimax problem for complex spectra. *Applied Mathematics Letters*, 1(3):311–314, 1988.
- [44] E.L. Wachspress. Trail to a Lyapunov equation solver. *Computers & Mathematics with Applications*, 55(8):1653–1659, 2008.
- [45] D.A. Wilson. Optimum solution of model-reduction problem. *Proc. IEE*, 117(6):1161–1165, 1970.
- [46] D. Zigic, L.T. Watson, and C.A. Beattie. Contragredient transformations applied to the optimal projection equations. *Linear algebra and its applications*, 188:665–676, 1993.